



JP8248986

Biblio

esp@cenet

**PATTERN RECOGNITION METHOD**

Patent Number: JP8248986
Publication date: 1996-09-27
Inventor(s): TAKAHASHI SATOSHI; SAGAYAMA SHIGEKI
Applicant(s):: NIPPON TELEGR & TELEPH CORP <NTT>
Requested Patent: ☐ JP8248986
Application Number: JP19950052391 19950313
Priority Number(s):
IPC Classification: G10L3/00 ; G10L3/00
EC Classification:
Equivalents:

Abstract

PURPOSE: To effectively reduce the number of model parameters without degrading the recognition performance by reducing the number of the total parameters of the entire model through the use of one common parameter for similar characteristic parameters.

CONSTITUTION: Against the inputted vectors, the likelihood of the hidden Markov model (HMM), in which the output probability distribution of each state is expressed by a multi-dimensional continuous distribution, is computed and the method outputs the category, which expresses the model having a highest likelihood, as the recognition result. Moreover, when it is expressed by a single continuous probability distribution or a mixed continuous probability distribution in each state of the HMM, one dimensional continuous distribution existing in each dimension of the multi-dimensional continuous distribution, which constitutes the above distributions, has common parameters, which express the distributions, between the distribution that exists in each dimension of other multi-dimensional continuous distributions. Furthermore, the method is provided with the HMM in which the commonality relationship of these parameters is specified in an individual manner. Thus, the parameters are effectively learned and the computation cost of the output probability is reduced.

Data supplied from the esp@cenet database - I2

BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-248986

(43) 公開日 平成8年(1996)9月27日

(51) Int. Cl. ⁴	識別記号	庁内整理番号	P I	技術表示箇所
G 1 0 L 3/00	5 3 5		G 1 0 L 3/00	5 3 5
	5 2 1			5 2 1 C

特許請求 未請求 請求項の数 4 O L (全 8 頁)

(21) 出願番号 特願平7-52391

(22) 出願日 平成7年(1995)3月13日

(71) 出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72) 発明者 高橋 敏

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72) 発明者 嵯峨山 茂樹

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(74) 代理人 弁理士 草野 卓

(54) 発明の名称 パターン認識方法

(57) 要約

【目的】 認識性能を劣化させずにモデルを表現する正規分布の総数を削減する。

【構成】 各音声単位をそれぞれ表現するHMMの各状態の多次元正規分布における各次元に存在する正規分布中の、平均値及び分散値が共に類似するものを共通化する。

(2)

特開平8-248986

1

【特許請求の範囲】

【請求項1】 入力ベクトルに対し、各状態の出力確率分布を多次元の連続分布で表現した隠れマルコフモデルの尤度を計算し、最も尤度の高いモデルが表現するカテゴリを認識結果として出力するパターン認識方法において、

上記隠れマルコフモデルの各状態が単一連続確率分布あるいは混合連続確率分布で表現されているときに、これらを構成する多次元連続分布の各次元に存在する一次元の連続分布が、他の多次元連続分布の各次元に存在する分布との間でその分布を表現する共通のパラメータを持ち、そのパラメータの共通化の関係が各次元で個別に規定されている隠れマルコフモデルを具備することを特徴とするパターン認識方法。

【請求項2】 上記多次元連続分布は多次元正規分布であり、上記共通のパラメータは平均値及び分散値であることを特徴とする請求項1記載のパターン認識方法。

【請求項3】 共通化されたパラメータの数が次元によって異なっている隠れマルコフモデルを具備することを特徴とする請求項1又は2記載のパターン認識方法。

【請求項4】 上記隠れマルコフモデルを学習などにより修正する場合において、共通化されたパラメータを一部のモデルにおいて修正して、すべてのモデルに含まれるその共通化されたパラメータを連動修正することを経特徴とする請求項1又は2記載のパターン認識方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 この発明は、隠れマルコフモデル (Hidden Markov Model, 以下HMMと記す) を用いて、入力ベクトルに対する各モデルの尤度を求めてその入力ベクトルの認識を行うパターン認識方法に関する。

【0002】

【従来の技術】 確率、統計論に基づいてモデル化するHMM法は、音声、文字、図形等のパターン認識において有用な技術である。以下では、音声認識を例にHMM法を用いた従来技術について説明する。従来の音声認識技*

$$b_s(X_t) = \sum_{k \in K_s} W_k^s P_k^s(X_t) \quad (1)$$

【0006】 のように計算される。ここで、 W_k^s は状態 s に含まれる k 番目の多次元正規分布 k に対する重み係数を表わす。多次元正規分布 k に対する確率密度 P_k^s 、

$$P_k^s(X_t) = \frac{1}{\sqrt{(2\pi)^r |\Sigma_k^s|}} \exp \left(-\frac{(X_t - \mu_k^s)^T \Sigma_k^{s-1} (X_t - \mu_k^s)}{2} \right) \quad (2)$$

2

* 屋において、認識すべき音声を用いてモデル化しておく方法は、性能が高く、現在の主流になっている。このHMM法の詳細は例えば文献1 (中川聖一：確率モデルによる音声認識、電子情報通信学会) に示されている。図2に従来のHMMを用いた音声認識装置の構成例を示す。入力端子11から入力された音声は、A/D変換部12においてデジタル信号に変換される。そのデジタル信号から音声特徴パラメータ抽出部13において音声特徴パラメータを抽出する。予め、認識しようとする音声単位 (例えば音素、音節、単語) ごとに作成したHMMをHMMメモリ14から読み出し、モデル尤度計算部15において、入力音声に対する各モデルの尤度を計算する。最も大きな尤度を示すモデルが表現する音声単位を認識結果として認識結果出力部16より出力する。

【0003】 図3Aに、3状態のHMMの例を示す。この様なモデルを音声単位 (カテゴリ) ごとに作成する。各状態 S_1 から S_3 には、音声特徴パラメータの統計的な分布 D_1 から D_3 がそれぞれ付与される。例えば、これが音素モデルであるとする、第1状態は音素の始端付近、第2状態は中心付近、第3状態は終端付近の特徴量の統計的な分布を表現する。

【0004】 各状態の特徴量分布は、複雑な分布形状を表現するために、複数の連続確率分布 (以下、混合連続分布と記す) を用いて表現される場合が多い。連続確率分布には、様々な分布が考えられるが、正規分布が用いられることが多い。また、それぞれの正規分布は、特徴量と同じ次元数の多次元無相関正規分布で表現されることが多い。図3Bに、混合連続分布の例を示す。この図では平均値ベクトルが μ_1 、分散値が σ_1 の正規分布 $N(\mu_1, \sigma_1)$ と $N(\mu_2, \sigma_2)$ と $N(\mu_3, \sigma_3)$ との3つの正規分布で表現された場合である。時刻 t の入力特徴量ベクトル $X_t = (x_{t1}, x_{t2}, \dots, x_{tr})^T$ (P は総次元数) に対する混合連続分布HMMの状態 s の出力確率 $b_s(X_t)$ は、

【0005】

【数1】

【0007】

【数2】

(3)

特開平8-248986

【0008】のように計算される。ここで、 $\mu_{k,i}^s$ はk番目の多次元正規分布kに対する平均値ベクトル、 $\Sigma_{k,i}^s$ は同じく共分散行列を表わす。共分散行列が対角成分のみ、つまり対角共分散行列であるとする、 $P^s(X_i)$ の対数値は、

$$\log P_k^s(X_i) = \sum_{i=1}^P \left(-\frac{P}{2} \log 2\pi - \log \sigma_{k,i}^s - \frac{(x_{k,i} - \mu_{k,i}^s)^2}{2\sigma_{k,i}^s} \right) \quad (3)$$

【0010】と表わせる。ここで、 $\mu_{k,i}^s$ は状態sの第k番目の多次元正規分布の平均値ベクトルの第i次目の成分を、 $\sigma_{k,i}^s$ は、状態sの第k番目の多次元正規分布の共分散行列の第i次目の対角成分（分散値）を表わす。この計算を認識候補のモデルについて、入力音声の各時刻の特徴量ベクトルに対して行い、得られた対数尤度をもとに認識結果を出力する。

【0011】

【発明が解決しようとする課題】認識性能を高めるためには、音響モデルの表現能力を高める必要がある。このためにモデルパラメータ数を増やす必要がある。多数のモデルパラメータを学習するためには膨大なデータ量が必要となるが、現実には限られたデータしか集めることができないので、むやみにモデルパラメータ数を増やすことができない。多数のパラメータを含むモデルを少量のデータで学習すると、学習データに深く依存したモデルとなってしまい、認識時に学習データとわずかに異なるデータに対しては認識誤りを起こしてしまう。しかし、モデルパラメータ数が少ないと表現能力が低いために十分な認識性能が得られない。このように、モデルの精度と認識時の頑健性にはトレードオフの関係があり、より少数のモデルパラメータで、より精密なモデルを表現する必要があるという問題がある。

【0012】また、HMMに基づくパターン認識装置において、式(2)の出力確率の計算コストが最も高い。音声認識装置の典型的な例では、この計算が消費する時間は音声認識処理時間全体の45%から65%を消費する。実時間処理はヒューマンインタフェースの観点からも重要な課題であるにもかかわらず、現状の処理速度は十分満足できるものではないという問題がある。

【0013】そこで、この発明の目的は、モデルの表現能力を保ちながら、認識性能を劣化させずに、モデルパラメータ数を効果的に削減し、同じデータ量でも効率的にパラメータを学習することができ、かつ出力確率の計算コストが少なく、実時間処理が可能なパターン認識方法を提供することにある。

【0014】

【課題を解決するための手段】この発明によれば、HMMの異なるモデル間や状態間に存在する分布のパラメータ

であっても、類似した性質のものは1つのパラメータで共通に使用して、モデル全体の総パラメータ数を削減する。例えば、多次元正規確率分布を構成する各次元の正規分布を表現する（規定する）パラメータ（平均値、分散値）を、類似した他の正規分布のパラメータとの間で共通化することを特徴とする。

【0015】

【実施例】この発明の方法を、多次元連続分布が正規分布である場合を例にして説明する。今、システム中のHMMのある2つの状態に存在する多次元正規分布A、Bに着目し、これらを、 $N(\mu_1^A, \Sigma_1^A)$ 、 $N(\mu_1^B, \Sigma_1^B)$ と表す。ここで、 $N(\mu, \Sigma)$ は、平均値ベクトルが μ で、共分散行列が Σ の正規分布であることを表現する。特徴パラメータの次元数を4とし、共分散行列は対角成分のみとする、これらの分布は、

$$\begin{aligned} N(\mu_1^A, \Sigma_1^A) &= \{N(\mu_{1,1}^A, \sigma_{1,1}^A), N(\mu_{1,2}^A, \sigma_{1,2}^A), N(\mu_{1,3}^A, \sigma_{1,3}^A), N(\mu_{1,4}^A, \sigma_{1,4}^A)\} \\ N(\mu_1^B, \Sigma_1^B) &= \{N(\mu_{1,1}^B, \sigma_{1,1}^B), N(\mu_{1,2}^B, \sigma_{1,2}^B), N(\mu_{1,3}^B, \sigma_{1,3}^B), N(\mu_{1,4}^B, \sigma_{1,4}^B)\} \end{aligned}$$

と表すことができる。ここで、それぞれの2次元目の正規分布、 $N(\mu_{1,2}^A, \sigma_{1,2}^A)$ と $N(\mu_{1,2}^B, \sigma_{1,2}^B)$ とが類似しているとき、これらを共有化し、1つの分布 $N(\mu_{1,2}^C, \sigma_{1,2}^C)$ で代表させ、それぞれの多次元正規分布を、

$$\begin{aligned} N(\mu_1^A, \Sigma_1^A) &= \{N(\mu_{1,1}^A, \sigma_{1,1}^A), N(\mu_{1,2}^C, \sigma_{1,2}^C), N(\mu_{1,3}^A, \sigma_{1,3}^A), N(\mu_{1,4}^A, \sigma_{1,4}^A)\} \\ N(\mu_1^B, \Sigma_1^B) &= \{N(\mu_{1,1}^B, \sigma_{1,1}^B), N(\mu_{1,2}^C, \sigma_{1,2}^C), N(\mu_{1,3}^B, \sigma_{1,3}^B), N(\mu_{1,4}^B, \sigma_{1,4}^B)\} \end{aligned}$$

と置き換える。これによりパラメータ μ, σ の総数を16個から14個に減らすことができた。

【0016】次に、上記の例において、時刻iの入力ベクトル $X_i = (x_{1,i}, x_{2,i}, x_{3,i}, x_{4,i})$ に対し尤度を計算する場合に、各次元の分布が共有化されていることの利点を述べる。尤度を各々のモデルに対して計算するためには、式(3)の計算が必要になる。初め

50

(4)

特開平8-248986

5

に、分布Aの尤度を計算したとする。次に、分布Bの計算を行うが、2次元目の分布は共有化されているので再計算する必要はなく、分布Aに対する2次元目の計算結果を利用することができる。このように、共有化されたパラメータに対する計算は、いずれかのモデルで計算がなされると、他のモデルにおいては計算結果を利用することができ、計算量の削減が図れる。

【0017】実際のシステムでは、例えば図1Aに示す様な要素パラメータのインデックステーブルを設け、つまり多次元正規分布A、Bに存在する各次元の正規分布、 $N(\mu_{1,1}, \sigma_{1,1}^2)$ 、 $N(\mu_{1,2}, \sigma_{1,2}^2)$ 、 $N(\mu_{1,3}, \sigma_{1,3}^2)$ 、 $N(\mu_{1,4}, \sigma_{1,4}^2)$ 、 $N(\mu_{2,1}, \sigma_{2,1}^2)$ 、 $N(\mu_{2,2}, \sigma_{2,2}^2)$ 、 $N(\mu_{2,3}, \sigma_{2,3}^2)$ 、 $N(\mu_{2,4}, \sigma_{2,4}^2)$ 、 $N(\mu_{3,1}, \sigma_{3,1}^2)$ 、 $N(\mu_{3,2}, \sigma_{3,2}^2)$ 、 $N(\mu_{3,3}, \sigma_{3,3}^2)$ 、 $N(\mu_{3,4}, \sigma_{3,4}^2)$ に對し、インデックス1、～1、をそれぞれ付け、各多次元分布A、Bをそれぞれ1、～1、で記述する。すなわち、 $N(\mu_{1,1}, \sigma_{1,1}^2) = (1, 1, 1, 1)$ 、 $N(\mu_{1,2}, \sigma_{1,2}^2) = (1, 1, 1, 1)$ とする。更に、計算結果バッファを図1Bに示すように設け、入力ベクトルに対する各次元の分布の計算結果を要素パラメータ1、を参照して格納する。計算結果バッファは初期状態では例えば-1を設定しておく。

【0018】入力ベクトルX、との計算に当たっては、例えば分布A ($N(\mu_{1,1}, \sigma_{1,1}^2)$) から行い、その1番目の要素インデックスが1、であるから、計算結果のバッファ (図1B) のインデックス1、に対する記憶が-1か否かを調べ、-1の場合は、図1Aの要素のパラメータのインデックステーブルを参照して、分布Aのパラメータ $\mu_{1,1}, \sigma_{1,1}^2$ を読み出し演算をして、その結果を計算結果バッファのインデックス1、の箇所に格納する。次の要素のインデックス1、についても同様に計算し、その計算結果を計算結果バッファのインデックス1、の箇所に格納する。以下同様にする。分布Bについての計算を行う際に、その2番目のインデックス1、については、計算結果バッファの箇所は先の分布Aについての計算の際に実行された結果が格納され、-1とは異なる値となっており、よってこの格納されている値を用いる。

【0019】この様なパラメータの共有化による計算結果の共有は、モデル数、状態数、状態内の分布数、特徴パラメータの次元数によらず行うことができる。次に、分布パラメータを各次元により異なる共有化関係で組み例について述べる。音声の特徴量ベクトルは、各次元ごとにパラメータが持つ情報量が異なる。例えば、特徴量の1つであるケプストラムは、16次元ぐらいで表現されることが多いが、主に低い次元の要素により多くの情報量が含まれている。そこで、低い次元の分布に対しては、共通化するパラメータの数を少なくし、自由度を高くして、分布の表現能力を高めておく。一方、高い次元の分布に対しては、パラメータの共通化を積極的に行

6

い、類似した分布は共通化しておく。このような操作により、分布の総数が同じでも、各次元ごとに分布の数を不均一に配置することによって、より効率的な表現が実現できる。必ずしも次元ごとに分布数を不均一に配置するのではなく、分布数が同一なる次元があってもよいことは明らかである。多次元正規分布を共通化する方法が提案されているが、この場合、ベクトルを共通化するために、各次元は同じ共有化関係が結ばれている。よって、共有化後に各次元に存在する分布の数は同じであり、情報量の多い次元も少ない次元も同じ数の分布で表現されている。したがって、同じ数の分布を持つモデルであっても (分布数が同じモデルの場合、計算量も一緒になる)、この発明による共通化方法によれば情報量の多い次元の分布数を多くし、情報量の少ない次元の分布数を少なくすることにより、より性能の高いモデルを構築することができる。言い換えれば、同じ性能のモデルを、より少ない数の分布で実現することができ、計算量が削減できる。

【0020】HMMのような統計的手法では、パラメータ数と学習データ量には密接な関係がある。数多くの分布がモデルに含まれていると、推定すべきパラメータ数が増加し、大量の学習データが必要になる。学習データ量が少ないと、モデルの一般性がなくなる。この発明では、認識のために必要な情報は残しながら、分布数を効果的に減らしているため、少ない学習データ量で、高い認識性能を得ることができる。

【0021】また、音声認識では、ある話者が発声した音声を用いて、不特定話者用に作られた音響モデルをその話者に合うように適応化する話者適応が行われる。話者適応の実際の場面では、少量の適応音声データしか得られないことがしばしばである。少量のデータでモデルパラメータを学習すると、パラメータが各モデルで独立に設定されている場合、学習データに関係する一部のモデルパラメータしか適応化できない。この発明の共有化を行えば、一部のモデルパラメータが適応化された場合、それらを共有する他のモデルのパラメータも同時に適応化できる。

【0022】次に実施例を述べる。HMMにおける共有化には (1) 異なる音素環境が同一のモデルを共有するモデルレベルの共有化 (第1階層) (例えば、嵯峨山 “音素環境のクラスタリング” 日本音響学会昭和62年度秋季研究発表会講演論文1-5-15や、K-F Lee 他 “Large-vocabulary speaker-independent continuous speech recognition using HMM” Proceedings of 1988 International Conference on Acoustics, Speech and Signal Processing, pp 123-126.参照)、(2) 異なるモデルが同一の状態を共有する状態レベルの共有化 (第2階層) (例えば、廣見、他 “逐次状態分割法 (SSS) による隠れマルコフネットワークの自動生成”、日本音響学会平成3年秋季研究発表会講演論文2-5

— 13 参照)。(3)異なる状態が同一の多次元正規分布を共有する基底分布レベル共有化(第3階層)(例えば、X.D. Huang "Unified technique for vector quantization and hidden Markov modeling using semi-continuous models", Proceedings of 1989 International Conference on Acoustics, Speech and Signal Processing, pp.639-642. 参照)。(4)この発明による異なる多次元正規分布の同一の一次元正規分布(平均値と分散値)を共有する特徴量レベルの共有化(第4階層)がある。

【0023】ステップ1：第1階層、第2階層の共有を実現するために、状態逐次分割法（SSS）を採用する。1名の話者データを用いて、各状態を単一正規分布*

$$d(i, j) = \left[\frac{1}{P} \sum_{k=1}^P \frac{(\mu_{ik} - \mu_{jk})^2}{\sigma_{ik} \sigma_{jk}} \right]^{\frac{1}{2}} \quad (4)$$

【0026】ここで、 μ_i 、 σ_i はそれぞれ平均値、分散値を表す。Pは総次元数である。各クラスタの代表分布は、クラスタ内の分布によって共有される。よって、各状態2混合を保ちながら、総分布数は900個から700個に削減される(3階層共有モデル)。

ステップ4: 特徴量レベルの共有化では、各次元を異なる数の分布で表現する。ここでは、ステップ2で得られたモデルの各次元に存在する900個の分布の平均分布間距離を計算し、その距離の比をもとに各次元の分布数 m 、(p は次元)を決定した。距離尺度は式(4)と同じである。

【0027】ステップ5: ステップ3で得られたモデルをもとに、各次元で独立にm個の分布クラスタを生成し、共有関係を決定する(4階層共有モデル)。なお、すべてのクラスタリングには最遠近隣法(Furthest neighbor method)を用いた。また、3階層共有モデル4階層共有モデルとも、分布の共有化関係が決定した後に再学習を行った。

【0028】このようにして得られた各階層モデルの性能と認識時の計算時間を、音声認識実験と単語認識実験で評価した。実験には、ATR重要語5240単語セットと216単語セットの男性10名分を使用した。このうち、8名を学習用話者、2名を評価話者とした。モデルの学習には、学習用話者の5240単語の偶数番目から10、480単語を平均的に選び出したセット、およびすべての学習用話者の216単語セットを使用した。音声認識実験の評価には、奇数番目の単語セットから524単語を任意に選び出し使用した。単語認識実験では、5240単語の奇数番目の単語から1500語を任意に選び出し認識対象語集とし、200単語を実際に認識して評価した。音声カテゴリー数は26である。パラメータは16次のケプストラム、16次の Δ ケプストラム

*で表現した450状態のHMnetを作成する。

ステップ2：上記モデルの各状態を2混合分布化した後、多数話者のデータで学習し、不特定話者用HMM-e1を作成する（2階層共有モデル）。

〔0024〕ステップ3:多次元正規分布レベルの共有を実現するために、すべての分布(2混合×450状態=900分布)から700個の分布クラスタを生成する(実験では、参考のため256個、64個の場合も行った)。分布*i*と分布*j*の距離尺度は以下のように定義した。

[0025]

【数4】

ム. Δ パワー (全部で33次元) である。

【0029】ステップ4において、特徴量の各次元に割り当てられた分布数の例を図4に示す。図4は各次元、平均64分布(総数: 64×33 次元 = 2112分布)とした場合の結果である。ケブストラム、△ケブストラムとも、低い次元の成分の方が平均的な分布間距離が大きく、より多くの分布が割り当てられていることがわかる。

【0030】次に図5に、各モデルの構成、平均音素認識率、平均単語認識率、平均計算時間（CPU time）の比を示す。図5中の3階層共有モデル、4階層共有モデルはどれも、常に450状態2混合（900分布）の構成を保ちながら共有化を行っている。4階層共有モデルでは、各次元の分布数は不均一に割り当てられているので、図5中の数値は平均分布数を示す。総パラメータ数には、平均値、共分散値、分布の重み係数が考慮されている。計算時間は、単語認識において出力確率計算に費やしたCPU timeを2階層共有モデルの場合を1.0として比で表した。計算機はSUN SPARC1.0である。参考のため、3状態2混合の音素環境独立モデルの結果も示す。

【0031】3層層共有モデルは、基底分布の数を700個に減らしても性能は変わらずに計算時間を短縮できるが、この場合は各次元のすべてについて平均値、共分散値が否かに類似しているとしているとするものであるから256個、64個にまで数を減らすと、もともと900個の分布に対し無理な共有関係が結ばれ図5に示すように性能（認識率）が急激に低下する。これに対し、この発明の4層層共有モデルでは、各次元に効果的に分布が割り振られ、それらの組み合わせで700個の基底分布を表現し、これらが900個の分布を表現している。これらの階層化によって、効率的な表現がなされ、

(6)

特開平8-248986

9

10

分布数が256でも音声認識率が87.6、単語認識率が90.0もあり、3階層共有モデルの分布数700と同一結果となり、分布数を64に減らしても、音声認識率が86.4、単語認識率が89.3と、わずかしこ減少せず、3階層共有モデルの分布数64よりも著しく高い認識率が得られ、しかも計算時間を削減できる。

【0032】このような効果が得られたのは、各次元で独立に正規分布の共有化をし、この結果、図4に示したように、次元の多い情報量に応じて分布数が異なり、つまり情報量の少ない高次元は多く共有化され、分布数が少なく、情報量が多い低次元は共有化が少なく、分布数が多く、全体として少ない分布数でモデルと効率的に表現することができるからである。

【0033】この発明は、音声認識に限らず、HMMを用いた文字認識、図形認識など、HMMを使用するあらゆるパターン認識において使用することができる。

【0034】

【発明の効果】この発明によるモデルパラメータの共有化の効果は2点ある。1つは、モデルの学習効率を上げることができる点、1つは認識時の計算量を削減できる点である。一般に、モデルパラメータは各モデルで独立に設定され、各々のカテゴリーのデータを用いて学習される。しかし、性質の類似したパラメータを異なるモデル間で共有化すれば、共有化したパラメータについては両方のカテゴリーのデータを用いて学習できるので、見かけ上のデータ量は増えることになる。

【0035】例えば、音声環境依存モデルには多数の多次元正規分布（例えば1000個以上）が存在するので各次元には同数の一次元正規分布が存在する。例えば、正規分布が各次元で1000個から m 個にマージされた場合でも、 m 個の正規分布の組み合わせで多次元正規分布を表現するので、それらが表現できる多次元正規分布の数は m^P 個であり（ P は次元数）、共有化後もかなり*

*の表現能力を保持している。

【0036】次に、計算量の観点から利点を考える。現在の多くのHMMは、多次元無相関正規分布を仮定しているため、対数尤度は(3)式のように計算される。

(3)式は、各次元において、正規分布に対する確率密度値の和になっており、次元間に渡る計算はない。したがって、計算は各次元で独立に考えることができる。分布を共有すれば、(3)式の各次元の結果をモデル間で共有でき、認識時の計算量の削減が可能であるという利点がある。

【0037】なお各分布中の平均値のみを共通化することが提案されているが、例えば、33次元の特徴量を持つ多次元正規分布が全部で1000個ある場合を考える（音声認識で実用的な値）。パラメータ数は2（平均値、分散値） $\times 33 \times 1000 = 66000$ 個となる。これを平均値のみ64個まで共通化すると、 $33 \times 1064 = 35112$ となる。平均値と分散値を共有化すると $33 \times 64 = 2112$ 個となり、大幅に削減可能である。計算量は、平均値のみを共有化した場合より更にわり算を各1回省略することができる。

【図面の簡単な説明】

【図1】Aは多次元正規分布要素インデックステーブルの例を示す図。Bは計算結果バッファテーブルの例を示す図である。

【図2】HMMを用いたパターン認識装置の一般的構成を示すブロック図。

【図3】AはHMMの例を示す図、Bは混合分布の例を示す図である。

【図4】この発明において特徴量の各次元に割り振られた分布数の実験例を示す図。

【図5】従来の共有化法、この発明による共有化法の認識性能と計算時間の実験結果を示す図。

【図4】

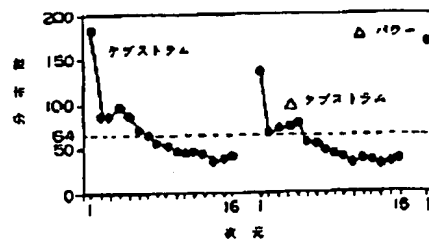


図 4

(7)

特開平8-248986

【図1】

図 1

	I_1	I_2	I_3	I_4	I_5	I_6	I_7
A	$N(\mu_{11}^1, \sigma_{11}^1)$	$N(\mu_{12}^1, \sigma_{12}^1)$	$N(\mu_{13}^1, \sigma_{13}^1)$	$N(\mu_{14}^1, \sigma_{14}^1)$	$N(\mu_{15}^1, \sigma_{15}^1)$	$N(\mu_{16}^1, \sigma_{16}^1)$	$N(\mu_{17}^1, \sigma_{17}^1)$
B	I_1	I_2	I_3	I_4	I_5	I_6	I_7
	-1	-1	-1	-1	-1	-1	-1

【図2】

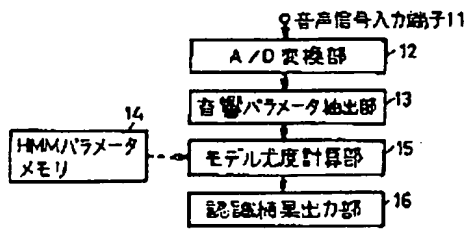
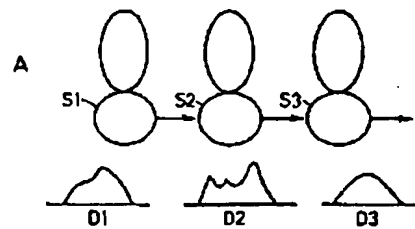


図 2

【図3】



B

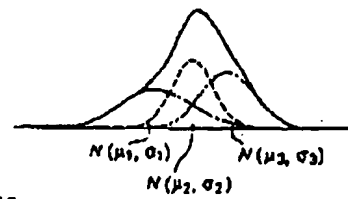


図 3

(8)

特開平8-248986

【図5】

【表5】

	状態数	組合数	最適化 分布数	各次元の 分布数	総パラ メータ数	学習効率 (%)	学習効率 (%)	計算時間 の比
各次元独立モデル	78	2	156	156	10452	78.4	72.5	0.17
2次元共有モデル	450	2	900	900	60300	87.7	90.0	1.00
		1	450	450	30150	86.0	86.0	0.49
3次元共有モデル		2	700	700	49000	87.6	90.0	0.81
			258	258	17152	86.7	87.5	0.35
			64	64	4228	77.2	77.0	0.15
4次元共有モデル		700	256	256	17696	87.6	90.0	0.70
			64	64	4224	86.4	89.9	0.48
			16	16	1758	84.2	89.5	0.43

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.